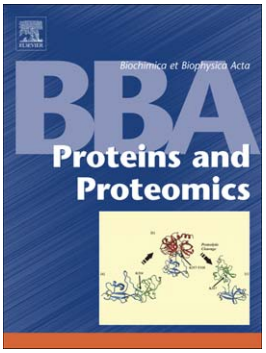
Bacterial and archaeal globins — A revised perspective

Vinogradov, S N ; Tinajero-Trejo, M ; Poole, R K ; Hoogewijs, D

Abstract: A bioinformatics survey of putative globins in over 2200 bacterial and some 140 archaeal genomes revealed that over half the bacterial and approximately one fifth of archaeal genomes contain genes encoding globins that were classified into three families: the M (myoglobin-like), and S (sensor) families all exhibiting the canonical 3/3 myoglobin fold, and the T family (truncated myoglobin fold). Although the M family comprises 2 subfamilies, flavohemoglobins (FHbs) and single domain globins (SDgbs), the S family encompasses chimeric globincoupled sensors (GCSs) and single domain Pgbs (protoglobins) and SSDgbs (sensor single domain globins). The T family comprises three classes TrHb1s,TrHb2s and TrHb3s, characterized by the abbreviated 2/2 myoglobin fold. The Archaea contain only Pgbs, GCSs and TrHb1s. The smallest globin-bearing genomes are the streamlined genomes (1.3 Mbp) of the SAR11 clade of alphaproteobacteria and the slightly larger (ca. 1.7 Mbp) genomes of Aquificae. The smallest genome with members of all three families is the 2.3 Mbp genome of the extremophile Methylacidiphilum infernorum (Verrumicrobia). Of the 147 possible combinations of the eight globin subfamilies, only 83 are observed. Although binary combinations are infrequent and ternary combinations are rare,the FHb+TrHb2 combination is the most commonly observed. Of the possible functions of bacterial globins we discuss the two principal ones - nitric oxide detoxification via the NO dioxygenase or denitrosylase activities and the sensing of oxygen concentration in the environmental niche. In only few cases a physiological role has been demonstrated in vivo.

DOI: https://doi.org/10.1016/j.bbapap.2013.03.021



Bacterial and Archaeal Globins – a Revised Perspective

Serge N. Vinogradov, Mariana Tinajero-Trejo, Robert K. Poole, David Hoogewijs

PII: S1570-9639(13)00134-9
DOI: doi: 10.1016/j.bbapap.2013.03.021
Reference: BBAPAP 39076

To appear in: *BBA - Proteins and Proteomics*

Received date: 14 December 2012
Revised date: 12 February 2013
Accepted date: 16 March 2013

Please cite this article as: Serge N. Vinogradov, Mariana Tinajero-Trejo, Robert K. Poole, David Hoogewijs, Bacterial and Archaeal Globins – a Revised Perspective, *BBA - Proteins and Proteomics* (2013), doi: 10.1016/j.bbapap.2013.03.021

Bacterial and Archaeal Globins – a Revised Perspective

Serge N. Vinogradov[a], Mariana Tinajero-Trejo[b], Robert K. Poole[b], David Hoogewijs*[c]

[a]Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, Michigan 48201, USA, [b]Institute of Biology and Biotechnology, Department of Molecular Biology and Biotechnology, The University of Sheffield, Sheffield, United Kingdom, [c]Institute of Physiology and Zürich Center for Integrative Human Physiology, University of Zürich, Zürich, Switzerland

Keywords: Bacteria, Archaea, Hemoglobin, Prokaryote

*Corresponding author:
Institute of Physiology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; Tel: +41-(0)44-635-5075; Fax: +41-(0)44-635-6814
Email address: david.hoogewijs@access.uzh.ch (David Hoogewijs)

Abstract

A bioinformatics survey of putative globins in over 2200 bacterial and some 140 archaeal genomes revealed that over half the bacterial and approximately one fifth of archaeal genomes contain genes encoding globins that were classified into three families: the M (myoglobin-like), and S (sensor) families all exhibiting the canonical 3/3 myoglobin fold, and the T family (truncated myoglobin fold). Although the M family comprises 2 subfamilies, flavohemoglobins (FHbs) and single domain globins (SDgbs), the S family encompasses chimeric globin-coupled sensors (GCSs) and single domain Pgbs (protoglobins) and SSDgbs (sensor single domain globins). The T family comprises three classes TrHb1s, TrHb2s and TrHb3s, characterized by the abbreviated 2/2 myoglobin fold. The Archaea contain only Pgbs, GCSs and TrHb1s. The smallest globin-bearing genomes are the streamlined genomes (~1.3 Mbp) of the SAR11 clade of alphaproteobacteria and the slightly larger (ca. 1.7 Mbp) genomes of Aquificae. The smallest genome with members of all three families is the 2.3 Mbp genome of the extremophile *Methylacidiphilum infernorum* (Verrumicrobia). Of the 147 possible combinations of the eight globin subfamilies, only 83 are observed. Although binary combinations are infrequent and ternary combinations are rare, the FHb+TrHb2 combination is the most commonly observed. Of the possible functions of bacterial globins we discuss the two principal ones - nitric oxide detoxification via the NO dioxygenase or denitrosylase activities and the sensing of oxygen concentration in the environmental niche. In only few cases a physiological role has been demonstrated *in vivo*.

1. Introduction

The first accounts of haemoglobins in microbes were those of Keilin in the 1930s, who demonstrated by use of a hand spectroscope, in which absorbance bands of a cell suspension are seen as black bands at characteristic wavelengths superimposed on a spectrum of visible colours [1-3]. This work lay fallow until B. Chance *et al*. in the 1970s investigated in biochemical detail the yeast protein and showed it to comprise not only a haem domain but also a flavin chromophore [4-5]; this was the first description of a flavohaemoglobin.

Recognition of haemoglobins in bacteria was slow to arrive. Dale Webster published several seminal papers in the 1970s describing a 'soluble cytochrome *o'* in the microaerobic bacterium *Vitreoscilla* [6-9]. The haem protein (variously named as Vgb, VHb etc.) was observed in intact cells as a stable oxygenated species [4-6]. The sequence of this protein [10] provided the first definitive proof that haemoglobins occur also in bacteria [11]. The first nucleotide sequence of a bacterial globin was that of Hmp in *E. coli*, the bacterial homologue of Keilin's and Chance's yeast protein. The sequence [12] clearly revealed the chimeric nature of the protein with an N-terminal globin domain and a C-terminal reductase domain with binding sites for NAD(P)H and FAD. More detailed accounts of the development of our now-extensive knowledge of bacterial flavohaemoglobins can be found elsewhere [13-14].

Concomitantly, a shorter than usual 118-residue globin, GlbN, was observed in *Nostoc commune* (Cyanobacteria: Nostocales) [15] and related cyanobacteria *Synechococcus* and *Synechocystis* [16]. Furthermore, a haemoprotein with kinase activity was found encoded by the oxygen sensor of *Rhizobium meliloti* [17]. Additional genomic information revealed the presence of single domain globins homologous to the globin domain of FHbs [14, 18] and of a family of

globin-coupled sensors involved in chemotaxis in Archaea and Bacteria [19-20] and related single domain protoglobins [21].

A survey of putative globins in some 250 bacterial and ca. 40 archaeal genomes showed that they all could be classified into one of three families and two structural classes [22-23]. The myoglobin-like family comprises chimeric (flavohemoglobin) and single domain globins as well as the sensor family comprising globin-coupled sensors (GCS), protoglobins (Pgbs), and sensor single domain globins (SSDgbs) all exhibiting the canonical secondary structure of Hbs, the 3/3 Mb-fold. The third family comprises the truncated globins consisting of three subgroups, all sharing an abbreviated 2/2 globin fold, with a vestigial or absent helix A and a loop substituting for helix F [24-26].

2. Material and Methods

2.1 Identification of globin sequences

The identification of putative globins and globin domains was carried out using two approaches. In one, we employed the globin gene assignments provided on the SUPERFAMILY site (http://supfam.mrc-lmb.cam.ac.uk), based on a library of hidden Markov models [27], discarded sequences shorter than 100 aa and subjected the remaining sequences to a FUGUE search [28] (www-cryst.bioc.cam.ac.uk). FUGUE scans a database of structural profiles, calculates the sequence-structure compatibility scores for each entry, using environment-specific substitution tables and structure-dependent gap penalties, and produces a list of potential homologs and alignments. FUGUE assesses the similarity between the query and a given structure via the Z score, the number of standard deviations above the mean score obtained by chance: the default threshold Z = 6.0 corresponds to 99% probability [28].

We also performed BLASTP and PSIBLAST searches with pairwise alignment [29] of the NCBI non-redundant protein sequence database (www.ncbi.nlm.nih.gov/BLAST/), using known globins as queries. PSIBLAST searches were carried out using the improved version with composition-based statistics [30]. The match between the query sequence and each of the sequences in a database is assessed by the Expect value (E), which is related to the probability of finding a false positive. Thus, E= 0.1 signifies that the probability of finding by chance, in a given database, another match with the query sequence having the same score, is 1 in 10. Iterations were carried out until all the sequences comprising the desired query group had E>0.005, the default threshold. Recognition is defined to be a hit with E>0.005, and with the pairwise alignment fulfilling the following two criteria: proper alignment of the F8 His residues and of helices BC through G.

2.2 Multiple sequence alignments and phylogenetic analysis

Multiple sequence alignments were carried out using MUSCLE [31], COBALT [32], MAFFT [33] PROBCONS [34] and TCOFFEE [35]. The quality of the alignments was assessed by MUMSA [36]. Neighor-joining (NJ) analyses were performed using MEGA version 5.05 [37]. Distances were corrected for superimposed events using the Poisson method. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option). The reliability of the branching pattern was tested by bootstrap analysis with 1000 replications. Bayesian inference trees were obtained employing MrBayes version 3.1.2 [38], assuming the WAG model of amino acid substitution and a gamma distribution of evolutionary rates, as determined by the substitution model testing option in MEGA 5.05. Two parallel runs, each consisting of four chains were run simultaneously for 8×10^6

generations and trees were sampled every 100 generations generating a total of 80000 trees. The final average standard deviations of split frequencies were stationary and posterior probabilities were estimated on the final 50000 trees. The CIPRES web portal was used for the Bayesian analyses [39] and MEGA version 5.05 [37] was used to visualize these analyses.

3. Results

3.1 Distribution of globins in bacterial and archaeal phyla

A list of >2000 bacterial genomes included in our survey is provided in Appendix A Supplementary Table A1. Globins were found in 1185 out of 2275 bacterial genomes (52%) and in 32 of 140 archaeal genomes (23%). Overall, all the globins found in this survey belong to one of the three families found earlier [22-23], along with the eight subfamilies. Table 1 lists the three families, the eight subfamilies and the estimated number of sequences; A small subset (1-5%) of sequences from each of the eight subfamilies lack the F8 His, the essential determinant of a globin, even though their FUGUE scores are >6. Fig.1 shows a Bayesian phylogenetic tree of about 150 manually selected representative globins, illustrating clearly the presence of three families and eight subfamilies.

A summary of the distribution of the eight globin subfamilies among the major bacterial and archaeal phyla is provided in Tables 2 and 3, respectively. The highest percent of globin-containing genomes are found in the Aquificae (100%) and the Deinococcus/Thermus group (88%). About 75% of alpha-, beta- and gamma-proteobacterial genomes and of planctomycete genomes have globins. Only about one half of the genomes in Cyanobacteria, delta- and epsilon-proteobacteria have globins. Fewer globins are found in Archaea (see Table 3) They are limited to Pgbs and SSDgbs in a few Crenarchaeota and to GCSs, Pgbs and TrHb1s in the Halobacteria (Euryarchaeota).The decrease in the

percent of globin-containing bacterial genomes from 65% to 53% found earlier [22] can be accounted for by the recent infusion of genomes from the Human Microbiome Project [40], the majority of which do not have globins.

3.2 How complete is the bacterial species coverage?

An appropriate question to ask at this juncture would be whether a new survey of globins in bacteria was necessary? We believe that it was, given that the earlier, limited survey [22-23], missed the presence of sensor single domain globins, the SSDgbs. At that time, their absence represented a substantial obstacle to the formulation of a model of globin evolution based on endosymbiotic events, particularly in view of the discovery of bacterial-like SSDgbs in fungi [41]. Thus, we feel justified in having carried out the new survey, to ensure that we would obtain the most complete coverage of bacterial globins possible. What about the bacterial species coverage? An evaluation of the approximately 1000 bacterial genomes available in 2007 [42] carried out by Wu *et al.* in 2009, concluded that the data suffered from a pronounced sampling bias, and suggested ways of correcting it [43]. Since then, more than 2000 completed genomes, including the Human Microbiome Project genomes [40] and those discussed in ref [43] have become available. Thus, although bias towards major bacterial groups such as Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria still exists, the coverage of bacterial species is probably adequate. A recent multiauthor study of the number of species on Earth and in the Oceans [44] has provided estimates of the currently known bacteria to be ca. 10,358 species on Earth and about 650 in the Oceans, with another predicted 9680 species; the number of Archaea is ca. 500 known on Earth, with another ca. 455 predicted. Given the >2000 bacterial and ca. 150 archaeal genomes employed in the present survey, it appears likely that we now have a fairly complete census of prokaryote globin diversity.

3.3 The smallest globin-containing genomes

Table 4 lists some of the smallest bacterial globin-containing genomes. These include the ca. 1.3 Mbp genomes of the SAR11 clade of alpha-proteobacteria, including *Candidatus Pelagibacter ubique*, one of the most widespread marine species and the slightly larger genomes of the Aquificiae. The smallest genome to contain members from all three globin families is *Methylacidiphilum infernorum* (Verrumicrobia), at 2.74 Mbp.

3.4 Correlation of genome size with globin content

There is a strong correlation between number of genes and genome size in bacteria, illustrated by a log-log plot in Fig. 2 borrowed from Giovannoni *et al.* [45]. Bacteria smaller than *Pelagibacter ubique,* a genome that exemplifies the so-called “streamlined” genome of free-living bacteria that have retained only the genes necessary for survival, include intracellular obligate symbionts wih highly reduced genomes that do not have globins (shown as black-filled circles). Although free-living bacteria (shown as green-filled circles in Fig. 2), generally do have globins, as exemplified by *Rhodopirellula baltica* and *Silicibacter pomeroy*, many do not, e.g. the cyanobacterial *Prochlorococcus* species. Likewise, among host-associated bacteria with genomes >4Mbp (shown by empty circles in Fig. 2) only some have globins. For example, although the *Streptomyces* genomes have similar sizes, and the majority have globins, several do not (see Appendix A Supplementary Table A1). Generally, globins are not found in the smaller genomes. Overall there also appears to be a linear relationship between genome size and number of globins, as exemplified by the plot for Actinobacteria shown in Fig. 3, in accord with the linear log-log plot shown in Fig. 2 [45]. These findings

emphasize that correlating globin function with bacterial lifestyle is fraught with difficulty.

3.5 Molecular phylogeny

We conducted phylogenetic analyses in two stages. In the first instance, NJ phylogenetic trees were generated of all globins identified in this study based on MAFFT multiple sequence alignments (L-INS-i option) for all 8 subfamilies separately. This approach allowed us to produce a strongly reduced set of globins with representative sequences of all 8 globin subfamilies. This set of 162 globin sequences comprising 53 S, 67 M and 42 T globins was used to generate a Bayesian phylogenetic tree, shown in Fig. 1. Two stress-responsive gene regulatory protein RsbR sequences, non-haem containing globin fold displaying sequences, from *Bacillus subtilis* and *Bacillus amyloliquefaciens* [46] were employed as outgroup. This global phylogenetic tree illustrates the presence of 3 distinct globin families, 2 of them supported with very high Bayesian probability values. Although the T family itself is not resolved, the T subfamilies (TrHb1, 2 and 3) clearly group separately as individual well-supported clusters. It should be noted that additional phylogenetic analyses with different outgroup sequences convincingly resolved the T family with highest Bayesian support (data not shown). Furthermore, within the S family, the Pgbs form a strictly supported monophyletic group separate from the SSDgbs that are interleaved with the GCSs. Note that the difference between Pgb and SSDgb is purely operational, defined by BLASTP searches that exploit the differences in their respective sequences. In contrast, there is much less difference between the FHb and SDgb sequences within the M family.

3.6 Combinations of globins observed in Bacteria

Many bacterial genomes comprise more than a single globin. We provide in Appendix B Supplementary Table B1 the observed combinations of the 8 globin subfamilies in 1162 genomes. Of the 147 possible combinations of the eight subfamilies (7 x 3 x 7=147), only 83 are observed. The most prevalent are single members. However a histogram of the results shown in Fig. 4 reveals the unusually frequent occurrence of the binary combination TrHb2 + FHb. Might the function of the TrHb2 globins be intimately linked to the established NO-detoxifying function of the FHbs? Interestingly, in *C. jejuni*, a genetically well-characterised species, a truncated globin coexists with a single-domain M-type globin that has proven NO-detoxifying ability. Other double and ternary combinations are infrequent.

3.7 Bacterial globins and pathogenicity

Among the 2275 bacterial genomes presented in this study, 246 (11.3%) belong to bacteria pathogenic for humans, animals or plants. Only 128 (52%) of these genomes contain globin sequences; this percentage correlates with the distribution of globin sequences found in the general database. When the genomes of pathogenic bacteria were classified according to oxygen requirements, 87 (35%) correspond to aerobic microorganisms, among which the proportion of globin-containing genomes is remarkably high (68%). Interestingly, all of the pathogenic aerobic microorganisms lacking globin sequences are host-associated. However, no assumptions should be made regarding host association and the absence of globin sequences. For instance, the genome of *Moraxella catarrhalis* lacks globin sequences, while three species of *Bordetella* (*B. bronchiseptica, B. parapertussis and B. pertussis*), possess a combination of TrHbs and one FHb sequences. All of them are aerobes, host-associated

mammalian respiratory tract pathogens. What factors are related to these notable differences? One possibility might simply be that the *Moraxella* genome is small ~2.0 Mbp, while the genomes from *Bordetella* species are large (5.3, 4.7 and ~4 Mbp respectively). However, since these bacteria share presumably similar niches within the host during infection and pathogenesis, only extensive experimental approaches can reveal the physiological significance(s) of the presence/absence of these globins.

4. Discussion

4.1 Relationship of bacterial globins to eukaryote globins

We now know that some of the bacterial globin subfamilies also occur in eukaryotes [22-23]. In particular, TrHb1s and TrHb2s occur in ciliates and algae, TrHb2s in plants [47], and FHbs and SSDgbs are found in fungi [48]. A number of bacterial SDgbs are identified as vertebrate Ngbs in the GenBank, reflecting the close homology between eukaryote 3/3 globins and the bacterial SDgbs and FHbs [49]. The extremophile *Methylacidiphilum infernorum* mentioned in Table 4 as being one of the smallest known bacterial genomes with representatives of all three globin families grows optimally at pH 2,0 and 60º C, and has 4 globins: a TrHb2, a SDgb, a FHb and an SSDgb [50]. The crystal structure of its SDgb was recently found to be very similar to that of vertebrate Ngbs [50], providing experimental support for our proposal of a phylogenetic link between vertebrate Ngbs and bacterial SDgbs and FHbs in a proposed model of globin evolution [41]. We postulated this link to exist as the result of a horizontal SDgb gene transfer during the endosymbiotic event involving the Last Eukaryote Common Ancestor (LECA) and an alpha-proteobacterium responsible for the origin of mitochondria [41].

4.2 Globin nomenclature

It is appropriate to consider at this point the problem of globin nomenclature. The GenBank identifications of non-vertebrate globins is generally haphazard and inconsistent, in the absence of an accepted overall globin phylogeny. Although FHbs are usually identified by the presence of an oxidoreductase domain together with a globin domain, the homologous single-domain globins are sometimes identified as a neuroglobin or oxidoreductase, or left unidentified. Likewise, while the GCSs are identified by the partner domain, e.g. as a

diguanylate cyclase or a methyl-accepting chemotaxis protein (MCP), the homologous single-domain globins are not even identified as globins. Furthermore, the truncated globins are identified as “cyanoglobins” or “protozoan globins”, and the group 2 globins are identified as O bacterial globins.

Based on Table 1, we propose in Table 5 the following comprehensive and self-consistent nomenclature for all globins, including eukaryotes. Let the three families be M (for myoglobin-like globins), retaining FHb (for flavohaemoglobins) and using SDgb (for single domain hemoglobin), S (for sensor), retaining GCS for chimeric members and using Pgbs (for protoglobins) and SSDgbs (for sensor single domain globins related to the GCS globin domain) and finally T (for truncated Mb-fold globins). Since the truncated (T) family is characterized by the 2/2 Mb-fold, its members have also been called 2/2 globins. Although the label ”2/2 globin” has been advocated in lieu of “truncated”, experience has demonstrated the awkwardness of repeated use of a label such as “2/2 group 1 or 2 or 3”: using TrHb1, TrHb2 or TrHb3 is much simpler. The terms N, O and P have also been used for the 3 T groups, but there is a precedent for using the numbering and the justification of the letters is not obvious. Furthermore, the nomenclature of the fast growing number of unicellular eukaryote Hbs should also be in line with the one proposed for bacterial globins. Over the last decade, several new globins were discovered in vertebrate genomes by Burmester, Hankeln and their collaborators [51]. These include the ubiquitous neuroglobins (Ngb), cytoglobins (Cygb) and others with more restricted distribution, the avian eye-specific globin E (GbE), globin X (GbX) and globin Y (GbY), as well as the recently discovered chimeric androglobins (Adgb) [52]. These are all 3/3 globins as are almost all the invertebrate globins.

Because of the sequence similarity between vertebrate globins such as Mb, Ngb and Cygb on one hand and the bacterial SDgbs and FHbs on the other [49], we propose subsuming the FHb family into the M family (for myoglobin-like), which would therefore encompass all 3/3 eukaryote globins. Thus, all the vertebrate globins and the 3/3 globins in microbial eukaryotes and in metazoa, should be considered as M family globins.

4.3. Non-heme bacterial globins

Several globin-like proteins lacking heme have been discovered recently. These include the *Bacillus subtilis* stress response regulator RsbR, where the G and H helices bend inward, eliminating the heme-binding cavity [46]. A dimeric globin fold that does not bind heme, is adopted by the sensor domain of one of several histidine kinases that regulate *Bacillus anthracis* sporulation; instead, they form a hydrophobic tunnel and hydrophilic chamber that are occupied by fatty acid, which engages a conserved arginine and chloride ion via its carboxyl head group [53]. *In vivo*, these domains may therefore recognize alterations in fatty acid levels, pH and chloride ion concentration. It appears to be a member of a large collection of sensor histidine kinases not only in Bacilli but also in Chlorobi and Deltaproteobacteria.

4.4 Functions of bacterial globins

Since only about half of the bacterial genomes have globins, it is evident that globins are not always indispensable. A recent estimate of a bacterial pangenome, based on the analysis of over 573 genomes [54], showed the average bacterial genome to have a core of shared essential genes accounting for about 8% of the total, together with a set of accessory genes providing distinction between species and serotypes, accounting for 28% of the genome, with the remaining 64% comprising genes necessary for colonization and survival in specific niches. It would appear that the globin genes belong to the class of genes that assist survival in certain niches but are not required universally for survival. It is interesting to point out that the dispensable bacterial SDgb genes having been transferred to LECA, subsequently underwent a highly successful adaptive evolution concomitant with the advent of multicellularity, resulting in globins capable of binding oxygen reversibly and become indispensable as oxygen transporters, at least in crown vertebrates [55].

We now know that Hbs have functions other than oxygen storage and transport, including enzymatic and sensing functions [56]. Although we have speculated on the possible functions of ancient globins [22-23, 55], it is worth reiterating here the following two most plausible hypotheses. One concerns the protective role of FHbs against nitrosative stress [13]. The other suggested by E. van Holde (Oregon State University), emphasizes the potential utility and survival value of sensor globins in chemotaxis, providing the means for the host microorganism to respond to alteration in oxygen concentrations [55]. The role of FHb in protection against nitrosative stress is due to its NO dioxygenation (NOD) [57] or denitrosylation activity [58]. Although all Hbs apparently have NOD activity, its physiological relevance is not always clear [57, 59-60] and in only a few species has this function been proven. In contrast to our understanding of NOD activity

[57, 61], we know very little about the GCS function, except for their involvement in positive as well as negative aerotactic responses. Apparently, the Pgbs and SSDgbs can also function as sensors, when coupled to an appropriate regulator domain [62]. The recent finding that GCSs can synthesize the second messenger bis-(3'-5')-cyclic diguanosine monophosphate in bacteria [63] suggests the possibility of additional functions for GCSs.

We now discuss how the explosion of publications has impacted, or not, upon our understanding of globin function in bacteria and archaea.

The number of globin sequences is vast yet the volume of experimental data on which to base any meaningful prediction of the significance of globin distribution between physiological groups or genera of bacteria is minimal. Fig. 5 attempts to quantify the problem by comparing the number of globin-like sequences available for analysis with the experimental data to hand. These data fall into the following general categories: estimated number of bacterial and archaeal globin-like sequences; heterologous expression to determine function; regulation studies; crystal structures; gene mutation/complementation studies.

However, these approaches are not of equal merit in learning about function. The sequence here and in the following text is, in our opinion, of increasing merit and value; (i) prediction of function based on analogy with related, well-studied examples; (ii) description of biochemical attributes of the purified protein *in vitro*; (iii) expression of the protein in a more tractable host cell type and study of the physiological attributes of the transformed host; (iv) elucidation of patterns of regulation of the globin and the resulting inferences about function *in vivo*; (v) mutagenesis and deletion of function in the native organism and subsequent analysis of the mutant phenotype. By this judgment, the ‘gold standard’ is the last - exploration of globin function by mutagenesis and phenotypic characterization. Sadly, the frequency of published papers is in overwhelming favour of those

approaches at the start of our list (c. 1161 globin-containing genomes) over the analysis of mutants and phenotypes (c. 15 in our estimation) (Fig. 5).

4.5 Experimental approaches to function determination

4.5.1 Crystal structures

A number of bacterial globin structures have been solved and published. These structures represent the eight subfamilies, except for SSDgbs. We have the crystal structures of two FHbs, *Alcaligenes eutrophus* (now known as *Cupriavidis necator*) [64] and *E. coli* [65], and four SDgbs, *Vitreoscilla* [66], *Pseudomonas aeruginosa* (PDB:1tu9), *Campylobacter jejuni* [67] and *Methylacidiphilum infernorum* [50]. The S family is represented by one Pgb, from *Methanosarcina acetivorans* [68] and two GCSs, *Bacillus subtilis* [69] and *Geobacter sulfurreducens* [70]. Numerous structures represent the TrHb family. These include four TrHb1s, *Mycobacterium tuberculosis* [71], *Geobacillus stearothermophilus* [72], and the cyanobacteria *Synechocystis sp.PCC7002* [73-74] *and Synechococcus sp. PCC7002* [75]; three TrHb2s, *Mycobacterium tuberculosis* [76], *Bacillus subtilis* [77] and *Thermobifida fusca [78]* and two TrHb3s, *Campylobacter jejuni* [79] and *Helicobacter hepaticus* [80]. The structures provide invaluable information on the choice of residues that can be fruitfully studied by site-directed mutagenesis and are critical, of course for quantum mechanical modeling of ligand access and other sub-molecular properties. Furthermore, structures identify important and unexpected posttranslational modifications, such as the covalent linkage of haem to protein in the cyanobacterial TrHb1s from *Synechococcus sp.PCC7002* and *Synechocystis sp.PCC7002* [81]. However, in our view, the structures have not been informative, in isolation, about globin function. For example, we are no closer to discerning a function for the truncated globin Ctb of *C. jejuni* (TrHb3, also known

as trHbP) [79]. Nevertheless, biological function can be shown to be consistent with the structural features once a function has been established. For example, the striking proximity of the FAD and haem moieties in the FHbs [64] is consistent with the exceptionally fast internal electron transfer rates required for rapid turnover [82] of the FHb proteins as dioxygenases (or denitrosylases) when provided with NAD(P)H for oxidation by the flavin and NO and O_2, which bind only to the haem. If we look outside the microbial field, we find that, despite the availability of structures, our understanding of the functions of cytoglobin and neuroglobin, remains incomplete.

4.5.2 Heterologous expression

A potentially valuable approach to determining or at least proposing function is to express the globin in a convenient host and then test the properties of the globin-expressing host. Expression of globins in *E. coli* is, of course, an exceptionally common first step in protein purification but surprisingly few of these studies report on whether the globin-expressing host exhibits unusual properties. As an example, it was clear long ago that high-level expression of the *E. coli hmp* gene in *E. coli* results in poor growth and may lead to spontaneous mutation of the cloned *hmp* gene (M. Tinajero-Trejo, unpublished). The detrimental effects of *hmp* expression are particularly severe when the recipient is challenged with peroxynitrite [83], peroxide [84] or is used to infect macrophages that generate these ROS [84]. It is now clear that these effects are the result of the over-production by Hmp of superoxide and hence peroxide [85]. With the benefit of hindsight it might have been possible to interpret correctly the deleterious effects of globin expression as the propensity for ROS generation when NO is absent. Overexpression of cytoglobin in *E. coli* leads to an extensive production of catalase, a green *d*-type haem-containing enzyme, and to the incorporation of *d*-

type haem (instead of the red *b*-type haem) in the apo-cytoglobin. The result is distinct “pine-green” cultures [86]. It is reasonable to propose that, on high-level expression, the globin generates superoxide and hence peroxide (as does Hmp) that leads to up-regulation in the host of catalase.

There are several applications of an *hmp* mutant of *E. coli*, which is consequently hypersensitive to NO and RNS. Where complementation of this deficiency is observed, a reasonable conclusion might be that the expressed globin possesses these detoxification properties and thus complements in *trans* the *hmp* mutation. There is, however, an important caveat: the demonstration of function in a heterologous host does not prove that the globin has the same function in the native organism. For example, the requirement in the NO dioxygenase reaction for oxygen and an appropriate source of reducing equivalents may be met in *E. coli* but not in the native host. Great care is needed. Nevertheless, where convenient genetic manipulation of the native microbe is precluded, for example, this approach can offer a clue to function.

A particular concern arises where the globin is presumed to require interaction with a cognate reductase. This is thought to be the case for those single-domain or truncated globins whose function is electron transfer; what is the source of the electrons? Rarely have bacterial globins been co-expressed in a heterologous host, sharing the same cytoplasm with the native, cognate reductase originating from the same bacterium. The earliest account involves the well-studied *Vitreoscilla* SDgb (VHb) and a reductase [87-88] isolated from that bacterium by Webster and colleagues. The existence of a robust reductase activity in *Vitreoscilla* was evident from the fact that an oxy-ferrous species was detectable in intact cells and that sustained oxygen reduction could be demonstrated [8]. Indeed VHb was for many years considered to be a soluble oxidase [7, 9] until the purified protein was sequenced [10] and shown to be a globin.

In some studies, however, the recipient host is designed to test a predicted function. The most obvious example of this approach relates again to VHb, which has been expressed in many cell types (microbes, plants and animals) and diverse functions were attributed to it in those hosts, notably detoxification of NO and RNS. This extensive field has been reviewed recently [89-90]. Interestingly, we are unaware of any evidence for NO and RNS metabolism in *Vitreoscilla* itself.

4.5.3 The gold standard – bacterial genetics as a tool for understanding globin function

The definitive test of globin function is, in our view, mutagenesis *in vivo* and description of the mutant phenotype. Needless to say, the mutation should meet normal standards of good practice, including the avoidance of polar effects on downstream genes, the extension of deletions into neighbouring up- or down-stream genes and complementation tests to verify that restoration of the wild-type phenotype correlates with wild-type gene function. Several examples are given in Table 6. In all cases, the functions deduced from the mutagenesis approach have been confirmed by other means. For example, the NO-sensitive phenotype of *hmp* mutants is corroborated by (a) the demonstration that *hmp* expression is up-regulated by NO [91], (b) numerous biochemical experiments on the NO detoxification reactions [92-94], and is supported by structural information (see above). With this approach, the choice of phenotype tests is critical. In the past, these have sometimes been guided by other studies, as in the case of *hmp*, in which a clean mutant was for the first time tested with NO and RNS [95] based on prior characterisation of the purified protein and expression studies with an *hmp-lacZ* fusion [95]. It is, of course, more readily attainable in microbial systems

than in the cells of higher organisms although there are a few clear examples of globin gene knockdown in higher eukaryotes.

Conclusion

The present study has emphasized the broad presence of globins belonging to three families in prokaryotes. We have identified globins in 1185 out of 2275 available bacterial genomes and in 32 of 140 archaeal genomes. We conclude that the efforts expended in studying globins have not been translated into reliable information on what these proteins do in the biology of the organism. The available data fall into several discrete approaches. The focus of this contribution is to highlight the remarkable number of globin (gene) sequences now known and to attempt to discern, predict, or guess what function(s) they may fulfill, a process fraught with uncertainty and difficulty.

Abbreviations

Adgb - androglobin; Cygb - cytoglobin; FHb – flavohaemoglobin; GbE – globin E; GbX – globin X; GbY - globin Y; GCS – globin-coupled sensor; Hb – haemoglobin; HGT - horizontal gene transfer; LECA - Last Universal Eukaryote Common Ancestor; Mb – myoglobin; Ngb - neuroglobin; Pgb – protoglobin; SDgb – single domain 3/3 globin related to the N-terminal of FHbs; SSDgb – sensor single domain 3/3 globin related to the N-terminal of GCSs.

Figure Legends:

Fig. 1. Bayesian phylogenetic tree of a selected reduced set of prokaryote globin sequences, representing 53 S (red/orange), 67 M (blue) and 42 T (green) globin family sequences. All globin sequences are identified by the first three letter of the genus name and the first three letters of the species name, the number of residues, the abbreviated phylum and family name, as well as accession numbers. In the S family the nonglobin domain is labeled with a letter: D – diguanylate cyclase; M – methyl accepting chemotaxis domain; H – histidine kinase domain; S – STAS (Anti-sigma-factor antagonist) domain; U – unidentified domain. Numbers next to the nodes correspond to Bayesian posterior probabilities (>0.90).

Fig. 2. A log-log plot of bacterial genome size versus number of genes from Giovannoni et al. [45], with permission from the American Association for the Advancement of Science.

Fig. 3. Plot of Actinobacterial genome size versus the number of putative globins observed in each genome.

Fig. 4. Histogram of globin combinations present in globin containing bacterial genomes based on Supplementary Table B1 in Appendix B.

Fig. 5. Comparison of the estimated number of globin-like sequences found in the bacterial and archaeal genomic databases with the experimental data reported for individual globins. Base 10 logarithm of (A) estimated number of globin-like sequences, (B) number of globins heterologously expressed for studies of function, (C) number of globins in which regulation has been explored (D) number of globins for which structures have been solved, and (E), number of globins studied by mutation/complementation.

List of Tables:

Table 1. Summary and structural properties of the three bacterial globin families and eight subfamilies.

Fold	3/3 Mb-fold		3/3 Mb-fold			2/2 Mb-fold		
Family name	Myoglobin-like		Globin-coupled sensor			Truncated		
Subfamily name	FHb (chimeric)	SDgb (single-domain)	GCS (chimeric)	Pgb	SSDgb (single-domain)	TrHb1 N	TrHb2 O	TrHb3 P
Estimated number of sequences	533	221	420	40	35	235	622	334

Table 2. Summary of globins present in bacterial genomes[1].

Taxon	Genomes with globins / total	Globin families
Actinobacteria (high G+C, Gram +)	**208/278** 76%	
Acidimicrobidae	1/1	GCS
Actinobacteridae		
Actinomycetales	207/277 (75%)	
Actinomyceneae	3/11	Pgb
Catenulisporineae	1/1	TrHb2, FHb, FHb*
Corynebacterinae	73/79	TrHb1,TrHb2, TrHb3, FHb
Frankineae	10/10	TrHb1, TrHb2, TrHb3, FHb
Glycomycineae	1/1	TrHb2, FHb
Kineosporiineae	1/1	TrHb2
Micrococcineae	40/42	TrHb2, FHb, Pgb
Micromonosporineae	9/10	TrHb1, TrHb2, FHb, SDgb
Propionibacterineae	5/14	TrHb1, TrHb2, FHb
Pseudonocardineae	15/17	TrHb2, FHb, Pgb
Streptomycineae	40/47	TrHb2, TrHb3, FHb, FHb*
Streptosporangineae	5/5	TrHb1, TrHb2, FHb, Pgb
Bifidobacteriales	0/16	
Coriobacteridae	1/18	Pgb
Nittriliruptoridae	0/1	
Rubrobacteridae	3/3	TrHb1, TrHb2, TrHb3, Pgb, GCS
Unclassified Actinobacteria	1/1	TrHb2, FHb
Aquificae	**12/12** 100%	
Aquificaceae	6/6	TrHb1, SDgb, GCS
Desulfurobacteriaceae	2/2	GCS
Hydogenothermaceae	4/4	TrHb1, GCS
Armatimonadetes	**0/1**	
Bacteroidetes / Chlorobi (Green Sulfur Bacteria)	**78/200 39%**	
Bacteroidales	6/97	TrHb2, TrHb3, SDgb, FHb
Cytophagia	19/19	TrHb1, TrHb2, TrHb3, SDgb, FHb, Pgb, SSDgb

Flavobacteria	40/59	TrHb2, TrHb3, SDgb, FHb
Sphingobacteria	12/13	TrHb1, TrHb2, TrHb3, FHb, Pgb, SSDgb
Chlorobi	1/12	SDgb
Caldiserica	**0/1**	
Chlamydiae / Verrucomicrobia	**14/29** 48%	
Chlamydiaea	3/13	TrHb2, FHb
Lentisphaerae	1/2	TrHb1, TrHb2, SDgb, FHb
Verrucomicrobia	10/14	TrHb1, TrHb2, SDgb,SDgb*, FHb, SSDgb
Chloroflexi (Green Nonsulfur bacteria)	**13/20** 61%	
Anaerolineae	0/1	
Caldilineae	1/1	Pgb, SSDgb
Chloroflexi	7/7	TrHb2, Pgb, SSDgb
Dehalococcoidetes	0/6	
Ktedonobacteria	1/1	
Thermomicrobia	4/4	Pgb,GCS
Chrysiogenetes	**1/1**	GCS
Cyanobacteria (Blue green algae)	**27/55** 50%	
Chroococcales	14/36	TrHb1, SDgb, SSDgb
Gloebacterales	1/1	TrHb1, SDgb, SSDgb
Nostocales	4/7	TrHb1, FHb*,SDgb
Oscillatoriales	7/9	TrHb1, SDgb, SSDgb
Prochlorales	0/1	
Stigonometales	1/1	SDgb, SSDgb
Deferribacteres	**2/4** 50%	GCS
Deinococococcus/Thermus	**14/16** 88%	
Deinococci	7/7	TrHb1, TrHb2, TrHb3, SDgb, FHb
Thermales	7/9	TrHb3, SDgb, Pgb, SSDgb
Dictyoglomi	**0/2**	
Elusimicrobia	**0/3**	
Fibrobacteres/Acidobacteria	**2/10** 25%	
Acidobacteria	2/9	TrHb1, TrHb3
Fibrobacteres	0/1	
Firmicutes (Low G + C, Gram +)	**116/564** 21%	
Bacilli	92/230	
Bacillales	92/99	
Alicyclobacillaceae	2/2	TrHb1, TrHb2, FHb, GCS
Bacillaceae	49/49	TrHb1, TrHb2, FHb, GCS
Listeriaceae	1/8	FHb
Paenibacillaceae	16/16	TrHb1, TrHb2, FHb, GCS

Planococcaceae	4/4	TrHb2, FHb, GCS
Sporolactobacillaceae	1/1	GCS
Staphylococcaceae	13/13	TrHb2, FHb
Thermoactinomycetaceae	1/1	TrHb1, TrHb2, FHb
Bacillales incertae sedis	5/5	TrHb2, FHb,FHb*, GCS
Lactobacillales	0/132	
Clostridia	22/276	
Clostridiales	18/177	SDgb, FHb,GCS
Halanaerobiales	0/5	
Natranaerobiales	0/1	
Thermoanaerobacteriales	4/35	SSDgb, GCS
Erysipelotrichi	2/23	SDgb
Negativicutes	0/35	
Fusobacteria	**1/22**	SDgb
Gemmatimonadetes	**1/1**	TrHb2, Sdgb
Ignavibacteria	**0/2**	
Nitrospirae	**5/5**	TrHb1, SDgb, SSDgb, GCS
Planctomycetes	**10/13**	
Phycisphaerae	1/1	TrHb1, TrHb2
Planctomycetia	10/12	TrHb1, TrHb2, SDgb, FHb, SSDgb, GCS
Poribacteria	**0/1**	
Proteobacteria (Purple bacteria)	**666/912** 73%	
Alphaproteobacteria	**183/256** 72%	
Caulobacterales	10/10	TrHb1, TrHb2, TrHb3, SDgb, FHb, GCS
Magnetococcales	1/1	GCS
Parvularculales	1/1	TrHb1
Rhizobiales	73/81	TrHb1, TrHb2, TrHb3, SDgb, FHb, GCS
Rhodobacterales	46/54	TrHb1, TrHb2, TrHb3, SDgb, FHb, SSDgb, GCS
Rhodospirilalles	29/29	TrHb1, TrHb3, FHb, GCS
Rickettsiales	2/46	TrHb3, FHb
Sphingomonadales	15/24	TrHb2, TrHb3, SDgb, FHb, GCS
Unclassified Alphaproteobacteria	6/10	
SAR11 cluster	2/4	TrHb1
SAR116 cluster	4/6	TrHb1, TrHb3, SDgb, FHb
Betaproteobacteria	**121/144** 85%	
Burkholderiales	84/93	TrHb1, TrHb2, TrHb3, FHb, GCS
Gallionellales	2/2	TrHb1, TrHb2, GCS
Hydrogenophilales	1/1	TrHb2, FHb
Methylophilales	6/6	TrHb1, TrHb2, SDgb, FHb, GCS
Neisseriales	14/27	TrHb2, SDgb, FHb, GCS
Nitrosomonadales	5/5	TrHb2, TrHb3, Pgb

Rhodocyclales	7/7	TrHb1, TrHb2, TrHb3, FHb, GCS
Unclassified Betaproteobacteria	2/3	TrHb2
Gammaproteobacteria	**307/402** 75%	
Acidithiobacillales	5/5	TrHb1, TrHb2, TrHb3, SDgb, GCS
Aeromonadales	6/8	FHb, GCS
Alteromonadales	53/61	TrHb1, TrHb2, TrHb3, SDgb, FHb SSDgb, GCS
Cardiobacteriales	0/1	
Chromatiales	12/20	TrHb1, TrHb3, Pgb, GCS
Enterobacteriales	86/104	TrHb2, FHb, GCS
Legionellales	4/7	TrHb1, TrHb3
Methylococcales	6/6	TrHb1, TrHb3, GCS
Oceanospirillales	15/17	TrHb1, TrHb2, TrHb3, SDgb, FHb, GCS
Pasteurellales	2/21	TrHb3, SDgb, FHb
Pseudomonadales	33/52	TrHb1, TrHb2, SDgb, FHb, GCS
Salinisphaerales	1/1	TrHb1, FHb
Thiotrichales	10/15	TrHb1, TrHb2, SDgb, FHb, GCS
Vibrionales	39/41	TrHb1, TrHb2, SDgb, FHb, GCS
Xanthomonadales	21/23	TrHb1, TrHb3, SDgb, FHb, Pgb GCS
Unclassified gammaproteobacteria	14/20	TrHb1, TrHb2, TrHb3, SDgb, GCS
Deltaproteobacteria	**29/62** 50%	
Bdellovibrionales	2/2	TrHb2, TrHb3, FHb
Desulfarculales	1/1	GCS
Desulfobacterales	1/7	GCS
Desulfovibrionales	5/20	GCS
Desulfurellales	0/1	
Desulfuromonadales	9/11	TrHb1, GCS
Myxococcales	11/11	TrHb1, TrHb2, SDgb, FHb, SSDgb, GCS
Syntrophobacterales	0/4	
Unclassified deltaproteobacteria	1/2	GCS
SAR324 cluster	0/3	
Epsilonproteobacteria	**22/46** 50%	
Campylobacterales	19/40	TrHb2, TrHb3, SDgb
Nautiliales	0/2	
Unclassified epsilonproteobacteria	3/4	TrHb2, GCS
Zetaproteobacteria	**1/1**	SSDgb
Unclassified proteobacteria	**1/1**	GCS
Spirochaetes	**14/49** 27%	
Brachyspiraceae	4/4	TrHb3
Leptospiraceae	10/14	TrHb2
Spirochaetaceae	0/31	

Synergistetes	**0/13**	
Tenericutes	**0/35**	
Thermodesulfobacteria	**0/2**	
Thermotogae	**0/15**	
Candidate Division NC10	**0/1**	
Candidate Division TG1	**0/2**	
Candidate Division WWE1	**0/1**	
Unclassified Bacteria	**3/6**	Pgb, GCS
Globins present/total	**1185/2275** **52%**	

[1]An asterisk indicate missing a F8His in a putative globin with a FUGUE Z score >6,0.

Table 3. Summary of globins present in archaeal Genomes[1].

.

Taxon	Genomes with globins / total	Globin families
	6/40 15%	
Crenarcheaeota		
Thermoprotei		
Acidilobales	0/1	
Desulfococcales	1/7	Pgb
Sulfolobales	0/6	
Thermoproteales	4/25	Pgb, SSDgb
Candidate Division OP1	1/1	Pgb
Euryarchaeota	**26/100** 23%	
Archaeoglobi	0/4	
Halobacteria	19/21	TrHb1, MCP, SSDgb
Methanobacteria	0/13	
Methanococci	3/14	SSDgb*
Methanomicrobia	2/19	Pgb
Mrthanopyri	0/1	
Nanohaloarchaea	0/1	
Thermococci	0/15	
Thermoplasmata	0/4	
Korarchaeota	0/1	
Nanorchaeota	0/1	
Thaumarchaeota	1/5	Pgb
Other Archaea	1/1	TrHb1
	32/140 23%	

[1]An asterisk indicate missing a F8His in a putative globin with a FUGUE Z score >6,0.

Table 4. List of the smallest globin-containing genomes.

Species	Taxon	Genome size, Mbp	Globins
Alpha proteobacterium HIMB114	Alphaproteobacteria (SAR11 cluster)	1.30	TrHb1
Candidatus Pelagibacter ubique		1.31	TrHb1
Methylophilales bacterium HTCC2181	Betaproteobacteria (Methylophilales)	1.30	TrHb1
Candidatus Aquiluna sp. IMCC133023	Actinobacteria (Micrococcinae)	1.36	TrHb2
Candidatus Rickettsia amblyommii	Alphaproteobacteria (Rickettsiales)	1.48	FHb
Aquifex aeolicus	Aquificae	1.59	SDgb
Thermocrinis albus	(Aquificaceae)	1.50	SDgb
Hydrogenobaculum sp. Y04AAS1		1.59	SSDgb
Hydrogenobacter thermophilus		1.74	TrHb1, SDgb, SSDgb
Candidatus Arthromitus sp. SFB	Firmicutes (Clostridia)	1.5 - 1.6	SDgb, FHb
Campylobacter jejuni	Epsilonproteobacteria (Campylobacterales)	1.64	TrHb3, SDgb
Gemella morbillorum	Firmicute (Bacilli)	1.75	TrHb2
Neisseria cinerea	Betaproteobacteria (Neisseriales)	1.87	TrHb2
Thermus thermophilus	Deinococcus/Thermus (Thermus)	1.86	SSDgb
Thermodesulfobium narugense	Firmicutes (Thermoanaerobacterales)	1.90	2MCP, DGC
Zymomonas mobilis	Alphaproteobacteria (Sphingomonadales)	1.9 - 2.1	2MCP
*Methylacidiphilum infernorum**	Verrumicrobia	2.30	TrHb2, SDgb, FHb, SSDgb

Table 5. Proposed global globin nomenclature.

Family name	M (myoglobin-like globins)		S (sensor globins)			T (truncated globins)		
Fold of globin domain	3/3		3/3			2/2		
Bacterial Subfamily name	FHb (flavohaemoglobins)	SDgb (single-domain globins related to FHbs)	GCS (globin-coupled sensors)	Pgb (protoglobins)	SSDgb (sensor single-domain globins)	TrHb1 N	TrHb2 O	TrHb3 P
Archaea	-	-	HemAT	Pgb	SSDgb	TrHb1	-	-
Eukaryotes	FHb	Mb, HbA, HbB, Cygb, GbX, GbY, GbE, Ngb, Adgb and all metazoan globins	-	-	-	TrHb1	TrHb2	-

Table 6. Bacterial haemoglobins in which physiological role(s) have been studied by gene mutation/complementation.

Bacterium	Haemoglobin sub-family	Globin name	Suggested function(s)	References
E. coli	FHb	Hmp	O_2 dependant NO detoxification to nitrate (alleviation of NO toxicity) Anaerobic reduction of NO to N_2O	[92, 95-106]
Salmonella enterica serovar *typhimurium*	FHb	Hmp	NO detoxification	[83, 107-112]
C. jejuni	SDgb	Cgb	NO detoxification	[113-114]
C. coli	SDgb	Cgb	NO detoxification	[113]
C. jejuni	TrHb (TrHb3)	Ctb	Oxygen metabolism	[115]
Ralstonia eutropha	FHb	FHPSt*	Accumulation of nitrous oxide during denitrification	[116]
Erwinia chrysanthemi	FHb	HmpX*	Implied in survival *in plant* and in synthesis of pectate lyases, NO detoxification	[117-118]
B. subtillis	FHb	HmpBs*	Anaerobic protection against prolonged nitrosative stress, NO detoxification	[119-120]
M. bovis	TrHb (TrHb1)	HbN	NO detoxification	[121]
Synechococcus sp.	TrHb (TrHb1)	GlbN	NO detoxification	[75]
Staphylococcus aureus	FHb	Hmp	NO detoxification under microaerophilic condition, resistance to azoles	[122-125]
Vibrio fischeri	FHb	Hmp	NO detoxification	[126]
Sinorhizobium meliloti	Putative FHb	-	NO detoxification	[127]
V. cholerae	FHb	HmpA	NO detoxification	[128-129]
P. aeruginosa	FHb	HmpPa*	NO detoxification	[130]
Pseudoalteromonas haloplanktis	TrHb (TrHb2)	PhHbO*	Oxidative and nitrosative stress resistance	[131]

* do not obey accepted norms for naming of bacterial gene products.

References

[1] D. Keilin, Haemoglobin in fungi; occurrence of haemoglobin in yeast and the supposed stabilization of the oxygenated cytochrome oxidase, Nature, 172 (1953) 390-393.

[2] D. Keilin, J.F. Ryley, Haemoglobin in Protozoa, Nature, 172 (1953) 451.

[3] D. Keilin, A. Tissieres, Haemoglobin in moulds: *Neurospora crassa* and *Penicillium notatum*, Nature, 172 (1953) 393-394.

[4] R. Oshino, T. Asakura, K. Takio, N. Oshino, B. Chance, B. Hagihara, Purification and molecular properties of yeast hemoglobin, Eur. J. Biochem., 39 (1973) 581-590.

[5] R. Oshino, N. Oshino, B. Chance, B. Hagihara, Studies on yeast hemoglobin - properties of yeast hemoglobin and its physiological function in cell, Eur. J. Biochem., 35 (1973) 23-33.

[6] Y. Orii, D.A. Webster, Oxygenated cytochrome o (*Vitreoscilla*) formed by treating oxidized cytochrome with superoxide anion, Plant Cell Physiol., 18 (1977) 521-526.

[7] D.A. Webster, Y. Orii, Oxygenated cytochrome o An active intermediate observed in whole cells of *Vitreoscilla*, J. Biol. Chem., 252 (1977) 1834-1836.

[8] D.A. Webster, Y. Orii, Physiological role of oxygenated cytochrome o: observations on whole-cell suspensions of *Vitreoscilla*, J. Bacteriol., 135 (1978) 62-67.

[9] D.A. Webster, Y. Orii, Biphasic recombination of photodissociated CO compound of cytochrome o (s) from *Vitreoscilla*, J. Biol. Chem., 260 (1985) 15526-15529.

[10] S. Wakabayashi, H. Matsubara, D.A. Webster, Primary sequence of a dimeric bacterial haemoglobin from *Vitreoscilla*, Nature, 322 (1986) 481-483.

[11] M.F. Perutz, A bacterial haemoglobin, Nature, 322 (1986) 405.

[12] S.G. Vasudevan, W.L. Armarego, D.C. Shaw, P.E. Lilley, N.E. Dixon, R.K. Poole, Isolation and nucleotide sequence of the hmp gene that encodes a haemoglobin-like protein in *Escherichia coli* K-12, Mol. Gen. Genet., 226 (1991) 49-58.

[13] M.T. Forrester, M.W. Foster, Protection from nitrosative stress: a central role for microbial flavohemoglobin, Free Radic. Biol. Med., 52 (2012) 1620-1633.

[14] G. Wu, L.M. Wainwright, R.K. Poole, Microbial globins, Adv. Microb. Physiol., 47 (2003) 255-310.

[15] M. Potts, S.V. Angeloni, R.E. Ebel, D. Bassam, Myoglobin in a cyanobacterium, Science, 256 (1992) 1690-1691.

[16] D.R. Hill, T.J. Belbin, M.V. Thorsteinsson, D. Bassam, S. Brass, A. Ernst, P. Boger, H. Paerl, M.E. Mulligan, M. Potts, GlbN (cyanoglobin) is a peripheral membrane protein that is restricted to certain *Nostoc* spp, J. Bacteriol., 178 (1996) 6587-6598.

[17] M.A. Gilles-Gonzalez, G.S. Ditta, D.R. Helinski, A haemoprotein with kinase activity encoded by the oxygen sensor of *Rhizobium meliloti*, Nature, 350 (1991) 170-172.

[18] J. Farres, M.P. Rechsteiner, S. Herold, A.D. Frey, P.T. Kallio, Ligand binding properties of bacterial hemoglobins and flavohemoglobins, Biochemistry, 44 (2005) 4125-4134.

[19] S. Hou, T. Freitas, R.W. Larsen, M. Piatibratov, V. Sivozhelezov, A. Yamamoto, E.A. Meleshkevitch, M. Zimmer, G.W. Ordal, M. Alam, Globin-coupled sensors: a class of heme-containing sensors in Archaea and Bacteria, Proc. Natl. Acad. Sci. U. S. A., 98 (2001) 9353-9358.

[20] S. Hou, R.W. Larsen, D. Boudko, C.W. Riley, E. Karatan, M. Zimmer, G.W. Ordal, M. Alam, Myoglobin-like aerotaxis transducers in Archaea and Bacteria, Nature, 403 (2000) 540-544.

[21] T.A. Freitas, S. Hou, E.M. Dioum, J.A. Saito, J. Newhouse, G. Gonzalez, M.A. Gilles-Gonzalez, M. Alam, Ancestral hemoglobins in Archaea, Proc. Natl. Acad. Sci. U. S. A., 101 (2004) 6675-6680.

[22] S.N. Vinogradov, D. Hoogewijs, X. Bailly, R. Arredondo-Peter, J. Gough, S. Dewilde, L. Moens, J.R. Vanfleteren, A phylogenomic profile of globins, BMC Evol. Biol., 6 (2006) 31.

[23] S.N. Vinogradov, D. Hoogewijs, X. Bailly, R. Arredondo-Peter, M. Guertin, J. Gough, S. Dewilde, L. Moens, J.R. Vanfleteren, Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life, Proc. Natl. Acad. Sci. U. S. A., 102 (2005) 11385-11389.

[24] A. Pesce, M. Couture, S. Dewilde, M. Guertin, K. Yamauchi, P. Ascenzi, L. Moens, M. Bolognesi, A novel two-over-two alpha-helical sandwich fold is characteristic of the truncated hemoglobin family, EMBO J., 19 (2000) 2424-2434.

[25] D.A. Vuletich, J.T. Lecomte, A phylogenetic and structural analysis of truncated hemoglobins, J. Mol. Evol., 62 (2006) 196-210.

[26] J.B. Wittenberg, M. Bolognesi, B.A. Wittenberg, M. Guertin, Truncated hemoglobins: a new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants, J. Biol. Chem., 277 (2002) 871-874.

[27] J. Gough, K. Karplus, R. Hughey, C. Chothia, Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure, J. Mol. Biol., 313 (2001) 903-919.

[28] J. Shi, T.L. Blundell, K. Mizuguchi, FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties, J. Mol. Biol., 310 (2001) 243-257.

[29] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res., 25 (1997) 3389-3402.

[30] A.A. Schaffer, L. Aravind, T.L. Madden, S. Shavirin, J.L. Spouge, Y.I. Wolf, E.V. Koonin, S.F. Altschul, Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements, Nucleic Acids Res., 29 (2001) 2994-3005.

[31] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res., 32 (2004) 1792-1797.

[32] J.S. Papadopoulos, R. Agarwala, COBALT: constraint-based alignment tool for multiple protein sequences, Bioinformatics, 23 (2007) 1073-1079.

[33] K. Katoh, K. Kuma, H. Toh, T. Miyata, MAFFT version 5: improvement in accuracy of multiple sequence alignment, Nucleic Acids Res., 33 (2005) 511-518.

[34] C.B. Do, M.S. Mahabhashyam, M. Brudno, S. Batzoglou, ProbCons: Probabilistic consistency-based multiple sequence alignment, Genome Res., 15 (2005) 330-340.

[35] P. Di Tommaso, S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J.M. Chang, J.F. Taly, C. Notredame, T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension, Nucleic Acids Res., 39 (2011) W13-17.

[36] T. Lassmann, E.L. Sonnhammer, Automatic assessment of alignment quality, Nucleic Acids Res., 33 (2005) 7120-7128.

[37] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol., 28 (2011) 2731-2739.

[38] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, Bioinformatics, 19 (2003) 1572-1574.

[39] M.A. Miller, W. Pfeiffer, T. Schwartz, Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, (2010) 1-8.

[40] X.C. Morgan, N. Segata, C. Huttenhower, Biodiversity and functional genomics in the human microbiome, Trends Genet., 29 (2013) 51-58.

[41] S.N. Vinogradov, D. Hoogewijs, X. Bailly, K. Mizuguchi, S. Dewilde, L. Moens, J.R. Vanfleteren, A model of globin evolution, Gene, 398 (2007) 132-142.

[42] K. Liolios, K. Mavromatis, N. Tavernarakis, N.C. Kyrpides, The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata, Nucleic Acids Res., 36 (2008) D475-479.

[43] D. Wu, P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N.N. Ivanova, V. Kunin, L. Goodwin, M. Wu, B.J. Tindall, S.D. Hooper, A. Pati, A. Lykidis, S. Spring, I.J. Anderson, P. D'Haeseleer, A. Zemla, M. Singer, A. Lapidus, M. Nolan, A. Copeland, C. Han, F. Chen, J.F. Cheng, S. Lucas, C. Kerfeld, E. Lang, S. Gronow, P. Chain, D. Bruce, E.M. Rubin, N.C. Kyrpides, H.P. Klenk, J.A. Eisen, A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea, Nature, 462 (2009) 1056-1060.

[44] C. Mora, D.P. Tittensor, S. Adl, A.G. Simpson, B. Worm, How many species are there on Earth and in the ocean?, PLoS Biol., 9 (2011) e1001127.

[45] S.J. Giovannoni, H.J. Tripp, S. Givan, M. Podar, K.L. Vergin, D. Baptista, L. Bibbs, J. Eads, T.H. Richardson, M. Noordewier, M.S. Rappe, J.M. Short, J.C. Carrington, E.J. Mathur, Genome streamlining in a cosmopolitan oceanic bacterium, Science, 309 (2005) 1242-1245.

[46] J.W. Murray, O. Delumeau, R.J. Lewis, Structure of a nonheme globin in environmental stress signaling, Proc. Natl. Acad. Sci. U. S. A., 102 (2005) 17320-17325.

[47] S.N. Vinogradov, I. Fernandez, D. Hoogewijs, R. Arredondo-Peter, Phylogenetic relationships of 3/3 and 2/2 hemoglobins in Archaeplastida genomes to bacterial and other eukaryote hemoglobins, Mol. Plant, 4 (2011) 42-58.

[48] D. Hoogewijs, S. Dewilde, A. Vierstraete, L. Moens, S.N. Vinogradov, A phylogenetic analysis of the globins in fungi, PLoS One, 7 (2012) e31856.

[49] S.N. Vinogradov, Tracing globin phylogeny using PSIBLAST searches based on groups of sequences, Methods Enzymol., 436 (2008) 571-583.

[50] A.H. Teh, J.A. Saito, A. Baharuddin, J.R. Tuckerman, J.S. Newhouse, M. Kanbe, E.I. Newhouse, R.A. Rahim, F. Favier, C. Didierjean, E.H. Sousa, M.B. Stott, P.F. Dunfield, G. Gonzalez, M.A. Gilles-Gonzalez, N. Najimudin, M. Alam, Hell's Gate globin I: an acid and thermostable bacterial hemoglobin resembling mammalian neuroglobin, FEBS Lett., 585 (2011) 3250-3258.

[51] T. Burmester, M. Haberkamp, S. Mitz, A. Roesner, M. Schmidt, B. Ebner, F. Gerlach, C. Fuchs, T. Hankeln, Neuroglobin and cytoglobin: genes, proteins and evolution, IUBMB Life, 56 (2004) 703-707.

[52] D. Hoogewijs, B. Ebner, F. Germani, F.G. Hoffmann, A. Fabrizius, L. Moens, T. Burmester, S. Dewilde, J.F. Storz, S.N. Vinogradov, T. Hankeln, Androglobin: A chimeric globin in metazoans that is preferentially expressed in Mammalian testes, Mol. Biol. Evol., 29 (2012) 1105-1114.

[53] G.R. Stranzl, E. Santelli, L.A. Bankston, C. La Clair, A. Bobkov, R. Schwarzenbacher, A. Godzik, M. Perego, M. Grynberg, R.C. Liddington, Structural insights into inhibition of *Bacillus anthracis* sporulation by a novel class of non-heme globin sensor domains, J. Biol. Chem., 286 (2011) 8448-8458.

[54] P. Lapierre, J.P. Gogarten, Estimating the size of the bacterial pan-genome, Trends Genet., 25 (2009) 107-110.

[55] S.N. Vinogradov, D. Hoogewijs, J.R. Vanfleteren, S. Dewilde, L. Moens, T. Hankeln, Evolution of the globin superfamily and its function, in: M. Nagai (Ed.) Hemoglobin: Recent Developments and Topics, Research Signpost, 2011, pp. 232-254.

[56] S.N. Vinogradov, L. Moens, Diversity of globin function: enzymatic, transport, storage, and sensing, J. Biol. Chem., 283 (2008) 8773-8777.

[57] P.R. Gardner, Hemoglobin, a nitric oxide dioxygenase, Scientifica, 1 (2012).

[58] A. Hausladen, A. Gow, J.S. Stamler, Flavohemoglobin denitrosylase catalyzes the reaction of a nitroxyl equivalent with molecular oxygen, Proc. Natl. Acad. Sci. U. S. A., 98 (2001) 10108-10112.

[59] P.R. Gardner, Nitric oxide dioxygenase function and mechanism of flavohemoglobin, hemoglobin, myoglobin and their associated reductases, J. Inorg. Biochem., 99 (2005) 247-266.

[60] B.J. Smagghe, J.T. Trent, 3rd, M.S. Hargrove, NO dioxygenase activity in hemoglobins is ubiquitous *in vitro*, but limited by reduction *in vivo*, PLoS One, 3 (2008) e2039.

[61] C.G. Mowat, B. Gazur, L.P. Campbell, S.K. Chapman, Flavin-containing heme enzymes, Arch. Biochem. Biophys., 493 (2010) 37-52.

[62] J.A. Saito, X. Wan, K.S. Lee, S. Hou, M. Alam, Globin-coupled sensors and protoglobins share a common signaling mechanism, FEBS Lett., 582 (2008) 1840-1846.

[63] X. Wan, J.R. Tuckerman, J.A. Saito, T.A. Freitas, J.S. Newhouse, J.R. Denery, M.Y. Galperin, G. Gonzalez, M.A. Gilles-Gonzalez, M. Alam, Globins synthesize the second messenger bis-(3'-5')-cyclic diguanosine monophosphate in bacteria, J. Mol. Biol., 388 (2009) 262-270.

[64] U. Ermler, R.A. Siddiqui, R. Cramm, B. Friedrich, Crystal structure of the flavohemoglobin from *Alcaligenes eutrophus* at 1.75 angstrom resolution, EMBO J., 14 (1995) 6067-6077.

[65] A. Ilari, A. Bonamore, A. Farina, K.A. Johnson, A. Boffi, The X-ray structure of ferric *Escherichia coli* flavohemoglobin reveals an unexpected geometry of the distal heme pocket, J. Biol. Chem., 277 (2002) 23725-23732.

[66] C. Tarricone, A. Galizzi, A. Coda, P. Ascenzi, M. Bolognesi, Unusual structure of the oxygen-binding site in the dimeric bacterial hemoglobin from *Vitreoscilla sp*, Structure, 5 (1997) 497-507.

[67] M. Shepherd, V. Barynin, C. Lu, P.V. Bernhardt, G. Wu, S.R. Yeh, T. Egawa, S.E. Sedelnikova, D.W. Rice, J.L. Wilson, R.K. Poole, The single-domain globin from the pathogenic bacterium *Campylobacter jejuni*: novel D-helix conformation, proximal hydrogen bonding that influences ligand binding, and peroxidase-like redox properties, J. Biol. Chem., 285 (2010) 12747-12754.

[68] M. Nardini, A. Pesce, L. Thijs, J.A. Saito, S. Dewilde, M. Alam, P. Ascenzi, M. Coletta, C. Ciaccio, L. Moens, M. Bolognesi, Archaeal protoglobin structure indicates new ligand diffusion paths and modulation of haem-reactivity, EMBO Rep., 9 (2008) 157-163.

[69] W. Zhang, G.N. Phillips, Jr., Structure of the oxygen sensor in *Bacillus subtilis*: signal transduction of chemotaxis by control of symmetry, Structure, 11 (2003) 1097-1110.

[70] A. Pesce, L. Thijs, M. Nardini, F. Desmet, L. Sisinni, L. Gourlay, A. Bolli, M. Coletta, S. Van Doorslaer, X. Wan, M. Alam, P. Ascenzi, L. Moens, M. Bolognesi, S. Dewilde, HisE11 and

HisF8 provide bis-histidyl heme hexa-coordination in the globin domain of *Geobacter sulfurreducens* globin-coupled sensor, J. Mol. Biol., 386 (2009) 246-260.

[71] M. Milani, A. Pesce, Y. Ouellet, P. Ascenzi, M. Guertin, M. Bolognesi, *Mycobacterium tuberculosis* hemoglobin N displays a protein tunnel suited for O_2 diffusion to the heme, EMBO J., 20 (2001) 3902-3909.

[72] A. Ilari, P. Kjelgaard, C. von Wachenfeldt, B. Catacchio, E. Chiancone, A. Boffi, Crystal structure and ligand binding properties of the truncated hemoglobin from *Geobacillus stearothermophilus*, Arch. Biochem. Biophys., 457 (2007) 85-94.

[73] C.J. Falzone, B. Christie Vu, N.L. Scott, J.T. Lecomte, The solution structure of the recombinant hemoglobin from the cyanobacterium *Synechocystis sp.* PCC 6803 in its hemichrome state, J. Mol. Biol., 324 (2002) 1015-1029.

[74] J.A. Hoy, S. Kundu, J.T. Trent, 3rd, S. Ramaswamy, M.S. Hargrove, The crystal structure of *Synechocystis* hemoglobin with a covalent heme linkage, J. Biol. Chem., 279 (2004) 16535-16542.

[75] N.L. Scott, Y. Xu, G. Shen, D.A. Vuletich, C.J. Falzone, Z. Li, M. Ludwig, M.P. Pond, M.R. Preimesberger, D.A. Bryant, J.T. Lecomte, Functional and structural characterization of the 2/2 hemoglobin from *Synechococcus* sp. PCC 7002, Biochemistry, 49 (2010) 7000-7011.

[76] M. Milani, P.Y. Savard, H. Ouellet, P. Ascenzi, M. Guertin, M. Bolognesi, A TyrCD1/TrpG8 hydrogen bond network and a TyrB10TyrCD1 covalent link shape the heme distal site of Mycobacterium tuberculosis hemoglobin O, Proc. Natl. Acad. Sci. U. S. A., 100 (2003) 5766-5771.

[77] L. Giangiacomo, A. Ilari, A. Boffi, V. Morea, E. Chiancone, The truncated oxygen-avid hemoglobin from *Bacillus subtilis*: X-ray structure and ligand binding properties, J. Biol. Chem., 280 (2005) 9192-9202.

[78] A. Bonamore, A. Ilari, L. Giangiacomo, A. Bellelli, V. Morea, A. Boffi, A novel thermostable hemoglobin from the actinobacterium *Thermobifida fusca*, FEBS J., 272 (2005) 4189-4201.

[79] M. Nardini, A. Pesce, M. Labarre, C. Richard, A. Bolli, P. Ascenzi, M. Guertin, M. Bolognesi, Structural determinants in the group III truncated hemoglobin from *Campylobacter jejuni*, J. Biol. Chem., 281 (2006) 37803-37812.

[80] H.J. Nothnagel, B.Y. Winer, D.A. Vuletich, M.P. Pond, J.T.J. Lecomte, Structural Properties of 2/2 Hemoglobins: The Group III Protein from *Helicobacter hepaticus*, IUBMB Life, 63 (2011) 197-205.

[81] M.R. Preimesberger, M.P. Pond, A. Majumdar, J.T. Lecomte, Electron self-exchange and self-amplified posttranslational modification in the hemoglobins from *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7002, J. Biol. Inorg. Chem., 17 (2012) 599-609.

[82] R.K. Poole, N. Ioannidis, Y. Orii, Reactions of the *Escherichia coli* flavohaemoglobin (Hmp) with oxygen and reduced nicotinamide adenine dinucleotide: evidence for oxygen switching of flavin oxidoreduction and a mechanism for oxygen sensing, Proc. Biol. Sci., 255 (1994) 251-258.

[83] S. McLean, L.A. Bowman, R.K. Poole, Peroxynitrite stress is exacerbated by flavohaemoglobin-derived oxidative stress in *Salmonella Typhimurium* and is relieved by nitric oxide, Microbiology, 156 (2010) 3556-3565.

[84] N.J. Gilberthorpe, M.E. Lee, T.M. Stevanin, R.C. Read, R.K. Poole, NsrR: a key regulator circumventing *Salmonella enterica* serovar Typhimurium oxidative and nitrosative stress in vitro and in IFN-gamma-stimulated J774.2 macrophages, Microbiology, 153 (2007) 1756-1771.

[85] G. Wu, H. Corker, Y. Orii, R.K. Poole, *Escherichia coli* Hmp, an "oxygen-binding flavohaemoprotein", produces superoxide anion and self-destructs, Arch. Microbiol., 182 (2004) 193-203.

[86] B.J. Reeder, D.A. Svistunenko, C.E. Cooper, M.T. Wilson, Engineering tyrosine-based electron flow pathways in proteins: the case of aplysia myoglobin, J. Am. Chem. Soc., 134 (2012) 7741-7749.

[87] V. GonzalesPrevatt, D.A. Webster, Purification and properties of NADH-cytochrome o reductase from *Vitreoscilla*, J. Biol. Chem., 255 (1980) 1478-1482.

[88] W. Jakob, D.A. Webster, P.M.H. Kroneck, NADH-dependent methemoglobin reductase from the obligate aerobe *Vitreoscilla* - Improved method of purification and reexamination of prosthetic groups, Arch. Biochem. Biophys., 292 (1992) 29-33.

[89] A.D. Frey, M. Shepherd, S. Jokipii-Lukkari, H. Haggman, P.T. Kallio, The single-domain globin of *Vitreoscilla*: augmentation of aerobic metabolism for biotechnological applications, Adv. Microb. Physiol., 58 (2011) 81-139.

[90] B.C. Stark, K.L. Dikshit, K.R. Pagilla, Recent advances in understanding the structure, function, and biotechnological usefulness of the hemoglobin from the bacterium *Vitreoscilla*, Biotechnol Lett, 33 (2011) 1705-1714.

[91] R.K. Poole, M.F. Anjum, J. Membrillo-Hernández, S.O. Kim, M.N. Hughes, V. Stewart, Nitric oxide, nitrite, and Fnr regulation of *hmp* (flavohemoglobin) gene expression in *Escherichia coli* K-12, J. Bacteriol., 178 (1996) 5487-5492.

[92] P.R. Gardner, A.M. Gardner, L.A. Martin, A.L. Salzman, Nitric oxide dioxygenase: an enzymic function for flavohemoglobin, Proc. Natl. Acad. Sci. U. S. A., 95 (1998) 10378-10383.

[93] A. Hausladen, A. Gow, J.S. Stamler, Flavohemoglobin denitrosylase catalyzes the reaction of a nitroxyl equivalent with molecular oxygen, Proc. Natl. Acad. Sci. U. S. A., 98 (2001) 10108-10112.

[94] C.E. Mills, S. Sedelnikova, B. SØballe, M.N. Hughes, R.K. Poole, *Escherichia coli* flavohaemoglobin (Hmp) with equistoichiometric FAD and haem contents has a low affinity for dioxygen in the absence or presence of nitric oxide, Biochem. J., 353 (2001) 207-213.

[95] J. Membrillo-Hernandez, M.D. Coopamah, M.F. Anjum, T.M. Stevanin, A. Kelly, M.N. Hughes, R.K. Poole, The flavohemoglobin of *Escherichia coli* confers resistance to a nitrosating agent, a "Nitric oxide Releaser," and paraquat and is essential for transcriptional responses to oxidative stress, J. Biol. Chem., 274 (1999) 748-754.

[96] A.M. Gardner, P.R. Gardner, Flavohemoglobin detoxifies nitric oxide in aerobic, but not anaerobic, *Escherichia coli*. Evidence for a novel inducible anaerobic nitric oxide-scavenging activity, J. Biol. Chem., 277 (2002) 8166-8171.

[97] P.R. Gardner, A.M. Gardner, L.A. Martin, A.L. Salzman, Nitric oxide dioxygenase: an enzymic function for flavohemoglobin, Proc. Natl. Acad. Sci. U. S. A., 95 (1998) 10378-10383.

[98] M.C. Justino, J.B. Vicente, M. Teixeira, L.M. Saraiva, New genes implicated in the protection of anaerobically grown *Escherichia coli* against nitric oxide, J. Biol. Chem., 280 (2005) 2636-2643.

[99] L. Svensson, M. Poljakovic, S. Save, N. Gilberthorpe, T. Schon, S. Strid, H. Corker, R.K. Poole, K. Persson, Role of flavohemoglobin in combating nitrosative stress in uropathogenic *Escherichia coli*--implications for urinary tract infection, Microb. Pathog., 49 (2010) 59-66.

[100] T.M. Stevanin, N. Ioannidis, C.E. Mills, S.O. Kim, M.N. Hughes, R.K. Poole, Flavohemoglobin Hmp affords inducible protection for *Escherichia coli* respiration, catalyzed by cytochromes bo' or bd, from nitric oxide, J. Biol. Chem., 275 (2000) 35868-35875.

[101] T.M. Stevanin, R.C. Read, R.K. Poole, The *hmp* gene encoding the NO-inducible flavohaemoglobin in *Escherichia coli* confers a protective advantage in resisting killing within macrophages, but not in vitro: links with swarming motility, Gene, 398 (2007) 62-68.

[102] H. Corker, R.K. Poole, Nitric oxide formation by *Escherichia coli.* Dependence on nitrite reductase, the NO-sensing regulator Fnr, and flavohemoglobin Hmp, J. Biol. Chem., 278 (2003) 31584-31592.

[103] P.R. Gardner, G. Costantino, A.L. Salzman, Constitutive and adaptive detoxification of nitric oxide in *Escherichia coli*. Role of nitric-oxide dioxygenase in the protection of aconitase, J. Biol. Chem., 273 (1998) 26528-26533.

[104] A. Hausladen, A.J. Gow, J.S. Stamler, Nitrosative stress: Metabolic pathway involving the flavohemoglobin, Proc. Natl. Acad. Sci. U. S. A., 95 (1998) 14100-14105.

[105] T.M. Stevanin, N. Ioannidis, C.E. Mills, S.O. Kim, M.N. Hughes, R.K. Poole, Flavohemoglobin Hmp affords inducible protection for *Escherichia coli* respiration, catalyzed by cytochromes *bo*' or *bd*, from nitric oxide, J. Biol. Chem., 275 (2000) 35868-35875.

[106] E. Hernandez-Urzua, C.E. Mills, G.P. White, M.L. Contreras-Zentella, E. Escamilla, S.G. Vasudevan, J. Membrillo-Hernandez, R.K. Poole, Flavohemoglobin Hmp, but not its individual domains, confers protection from respiratory inhibition by nitric oxide in *Escherichia coli*, J. Biol. Chem., 278 (2003) 34975-34982.

[107] N.J. Gilberthorpe, R.K. Poole, Nitric oxide homeostasis in *Salmonella typhimurium*: roles of respiratory nitrate reductase and flavohemoglobin, J. Biol. Chem., 283 (2008) 11146-11154.

[108] M.J. Crawford, D.E. Goldberg, Role for the *Salmonella* flavohemoglobin in protection from nitric oxide, J. Biol. Chem., 273 (1998) 12543-12547.

[109] T.M. Stevanin, R.K. Poole, E.A. Demoncheaux, R.C. Read, Flavohemoglobin Hmp protects *Salmonella enterica* serovar *typhimurium* from nitric oxide-related killing by human macrophages, Infect. Immun., 70 (2002) 4399-4405.

[110] I.S. Bang, L. Liu, A. Vazquez-Torres, M.L. Crouch, J.S. Stamler, F.C. Fang, Maintenance of nitric oxide and redox homeostasis by the *Salmonella* flavohemoglobin *hmp*, J. Biol. Chem., 281 (2006) 28039-28047.

[111] N.J. Gilberthorpe, M.E. Lee, T.M. Stevanin, R.C. Read, R.K. Poole, NsrR: a key regulator circumventing *Salmonella enterica* serovar *Typhimurium* oxidative and nitrosative stress in vitro and in IFN-gamma-stimulated J774.2 macrophages, Microbiology, 153 (2007) 1756-1771.

[112] Y.M. Park, H.J. Park, Y.H. Joung, I.S. Bang, Nitrosative stress causes amino acid auxotrophy in *hmp* mutant *Salmonella Typhimurium*, Microbiol. Immunol., (2011).

[113] K.T. Elvers, G. Wu, N.J. Gilberthorpe, R.K. Poole, S.F. Park, Role of an inducible single-domain hemoglobin in mediating resistance to nitric oxide and nitrosative stress in *Campylobacter jejuni* and *Campylobacter coli*, J. Bacteriol., 186 (2004) 5332-5341.

[114] M.S. Pittman, K.T. Elvers, L. Lee, M.A. Jones, R.K. Poole, S.F. Park, D.J. Kelly, Growth of *Campylobacter jejuni* on nitrate and nitrite: electron transport to NapA and NrfA via NrfH and distinct roles for NrfA and the globin Cgb in protection against nitrosative stress, Mol. Microbiol., 63 (2007) 575-590.

[115] L.M. Wainwright, K.T. Elvers, S.F. Park, R.K. Poole, A truncated haemoglobin implicated in oxygen metabolism by the microaerophilic food-borne pathogen *Campylobacter jejuni*, Microbiology, 151 (2005) 4079-4091.

[116] R. Cramm, R.A. Siddiqui, B. Friedrich, Primary sequence and evidence for a physiological function of the flavohemoprotein of *Alcaligenes eutrophus*, J. Biol. Chem., 269 (1994) 7349-7354.

[117] S. Favey, G. Labesse, V. Vouille, M. Boccara, Flavohaemoglobin HmpX: a new pathogenicity determinant in *Erwinia chrysanthemi* strain 3937, Microbiology, 141 (Pt 4) (1995) 863-871.

[118] M. Boccara, C.E. Mills, J. Zeier, C. Anzi, C. Lamb, R.K. Poole, M. Delledonne, Flavohaemoglobin HmpX from *Erwinia chrysanthemi* confers nitrosative stress tolerance and affects the plant hypersensitive reaction by intercepting nitric oxide produced by the host, Plant J., 43 (2005) 226-237.

[119] M.M. Nakano, Essential role of flavohemoglobin in long-term anaerobic survival of *Bacillus subtilis*, J. Bacteriol., 188 (2006) 6415-6418.

[120] A. Rogstam, J.T. Larsson, P. Kjelgaard, C. von Wachenfeldt, Mechanisms of adaptation to nitrosative stress in *Bacillus subtilis*, J. Bacteriol., 189 (2007) 3063-3071.

[121] H. Ouellet, Y. Ouellet, C. Richard, M. Labarre, B. Wittenberg, J. Wittenberg, M. Guertin, Truncated hemoglobin HbN protects *Mycobacterium bovis* from nitric oxide, Proc. Natl. Acad. Sci. U. S. A., 99 (2002) 5902-5907.

[122] A.R. Richardson, P.M. Dunman, F.C. Fang, The nitrosative stress response of *Staphylococcus aureus* is required for resistance to innate immunity, Mol. Microbiol., 61 (2006) 927-939.

[123] V.L. Goncalves, L.S. Nobre, J.B. Vicente, M. Teixeira, L.M. Saraiva, Flavohemoglobin requires microaerophilic conditions for nitrosative protection of *Staphylococcus aureus*, FEBS Lett., 580 (2006) 1817-1821.

[124] L.S. Nobre, V.L. Goncalves, L.M. Saraiva, Flavohemoglobin of *Staphylococcus aureus*, Methods Enzymol., 436 (2008) 203-216.

[125] L.S. Nobre, S. Todorovic, A.F. Tavares, E. Oldfield, P. Hildebrandt, M. Teixeira, L.M. Saraiva, Binding of azole antibiotics to *Staphylococcus aureus* flavohemoglobin increases intracellular oxidative stress, J. Bacteriol., 192 (2010) 1527-1533.

[126] Y. Wang, A.K. Dunn, J. Wilneff, M.J. McFall-Ngai, S. Spiro, E.G. Ruby, *Vibrio fischeri* flavohaemoglobin protects against nitric oxide during initiation of the squid-Vibrio symbiosis, Mol. Microbiol., 78 (2010) 903-915.

[127] E. Meilhoc, Y. Cam, A. Skapski, C. Bruand, The response to nitric oxide of the nitrogen-fixing symbiont *Sinorhizobium meliloti*, Mol. Plant. Microbe Interact., 23 (2010) 748-759.

[128] B.W. Davies, R.W. Bogard, N.M. Dupes, T.A. Gerstenfeld, L.A. Simmons, J.J. Mekalanos, DNA damage and reactive nitrogen species are barriers to *Vibrio cholerae* colonization of the infant mouse intestine, PLoS Pathog., 7 (2011) e1001295.

[129] A.M. Stern, A.J. Hay, Z. Liu, F.A. Desland, J. Zhang, Z. Zhong, J. Zhu, The NorR regulon is critical for *Vibrio cholerae* resistance to nitric oxide and sustained colonization of the intestines, MBio, 3 (2012) e00013-00012.

[130] H. Arai, M. Hayashi, A. Kuroi, M. Ishii, Y. Igarashi, Transcriptional regulation of the flavohemoglobin gene for aerobic nitric oxide detoxification by the second nitric oxide-responsive regulator of *Pseudomonas aeruginosa*, J. Bacteriol., 187 (2005) 3960-3968.

[131] E. Parrilli, M. Giuliani, D. Giordano, R. Russo, G. Marino, C. Verde, M.L. Tutino, The role of a 2-on-2 haemoglobin in oxidative and nitrosative stress resistance of Antarctic *Pseudoalteromonas haloplanktis* TAC125, Biochimie, 92 (2010) 1003-1009.

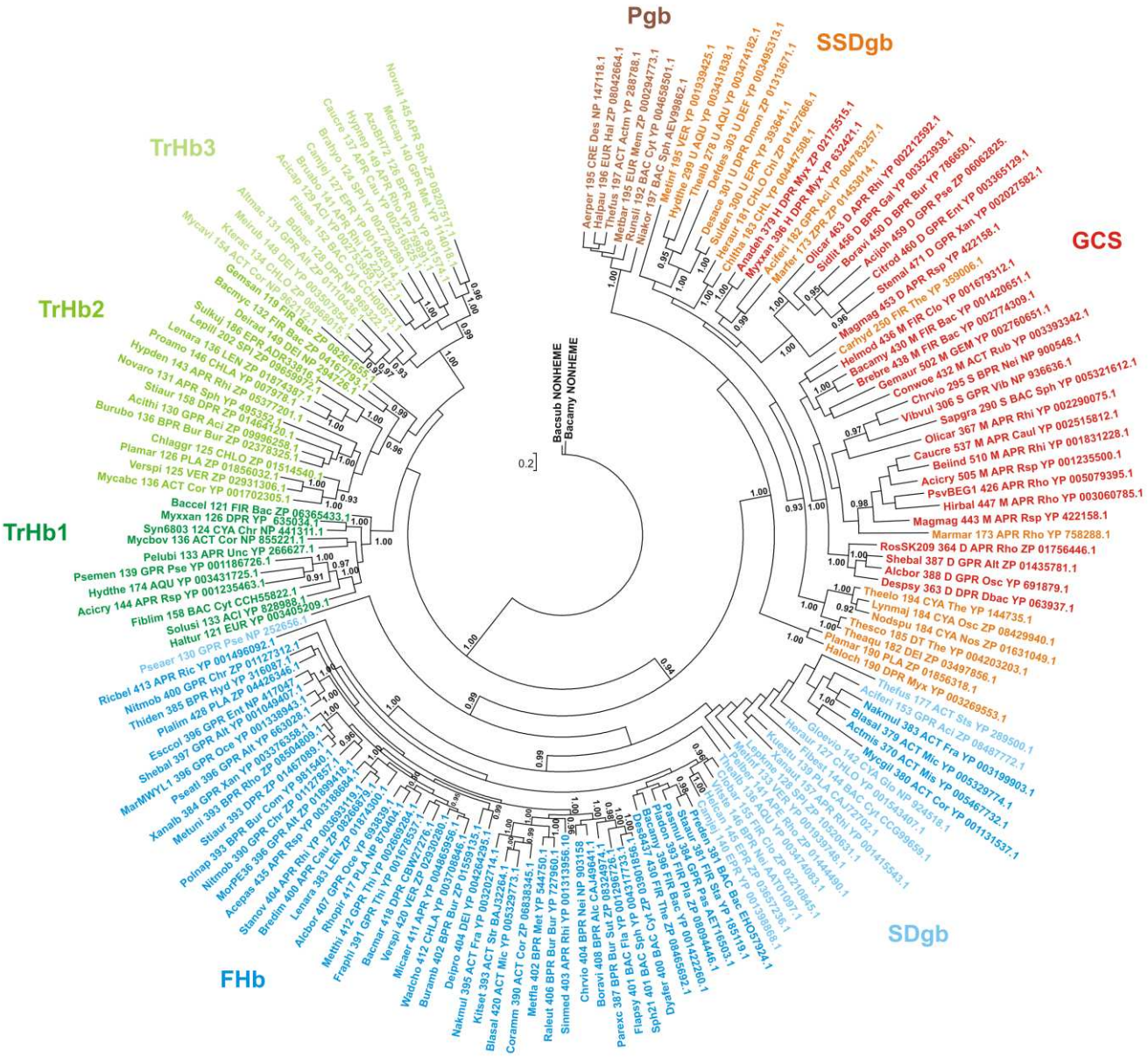


Figure 1

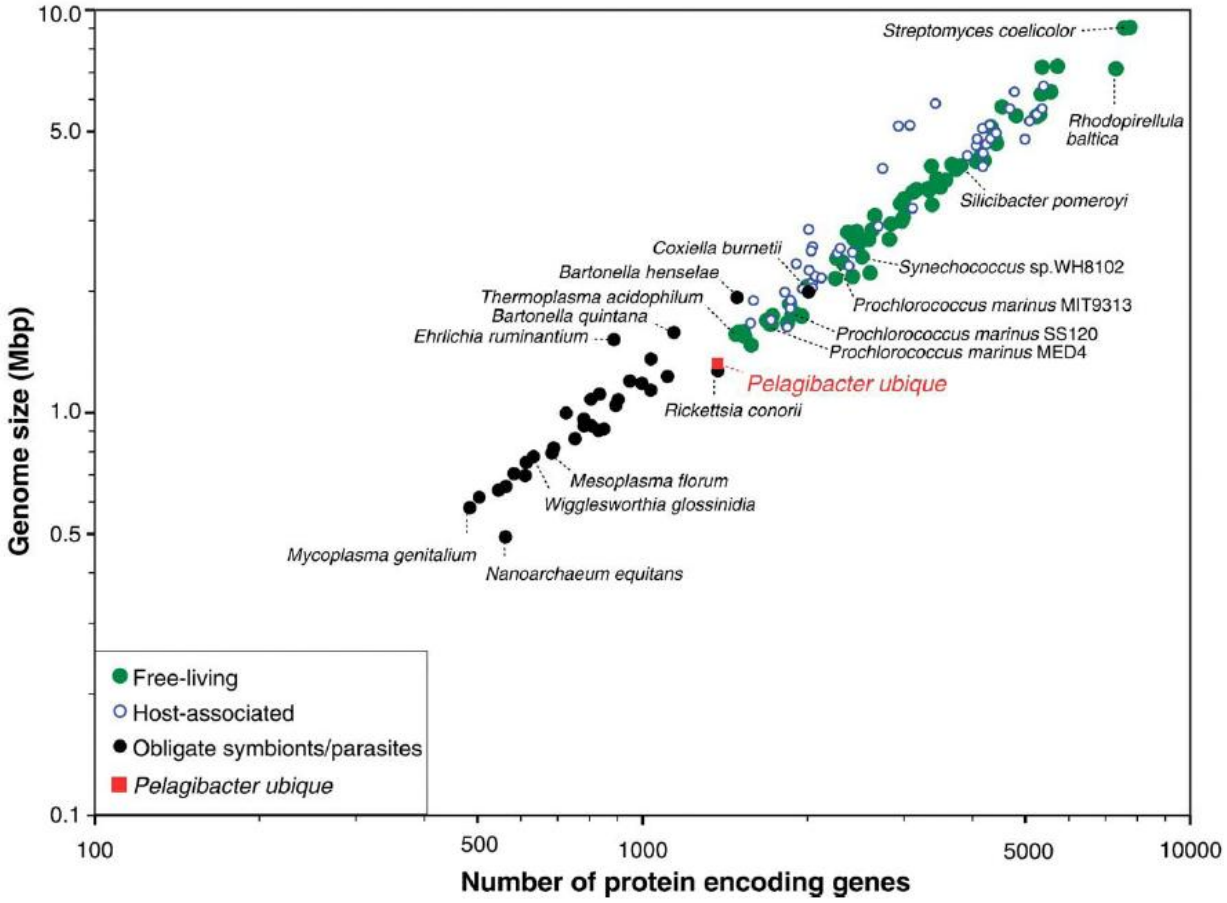


Figure 2

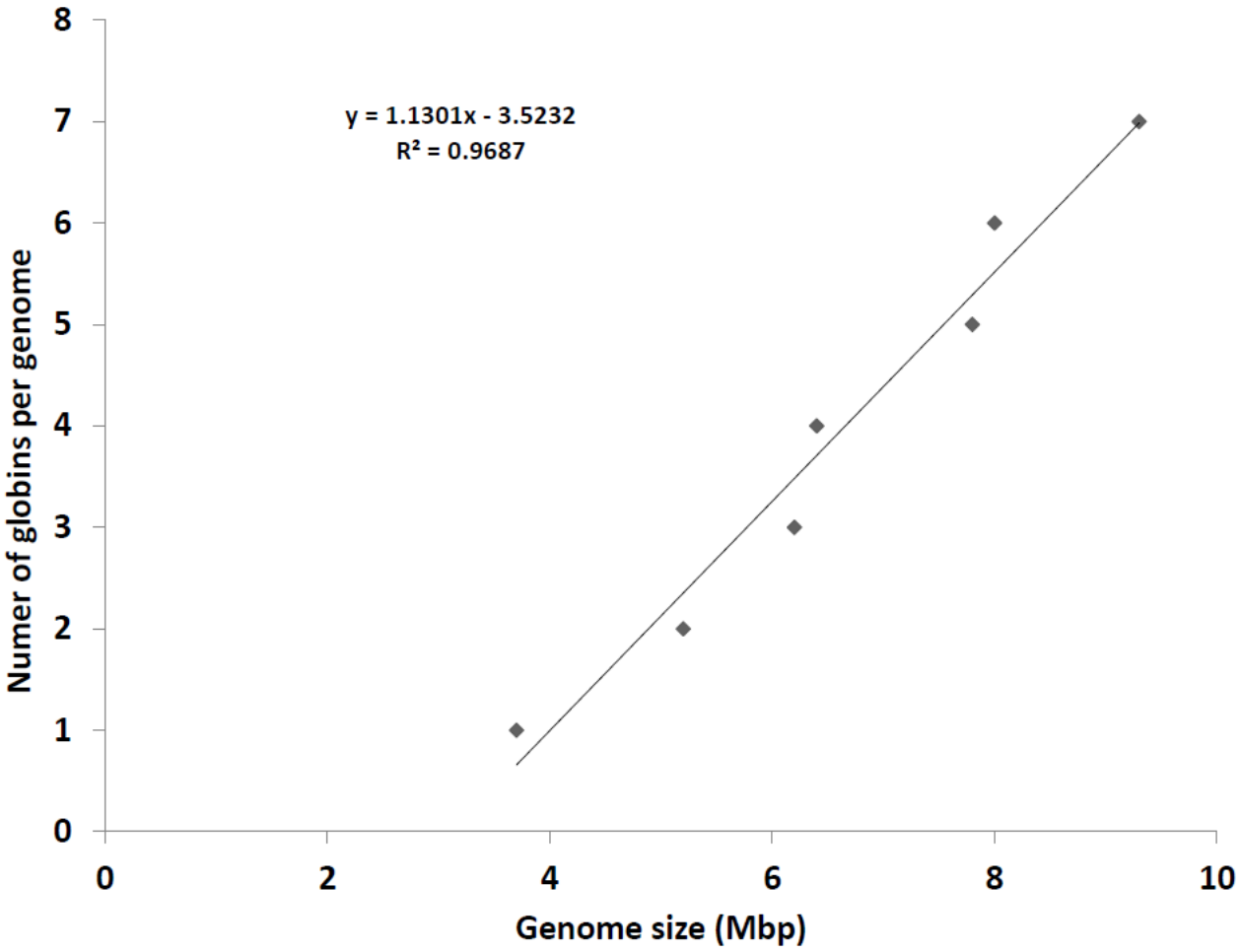


Figure 3

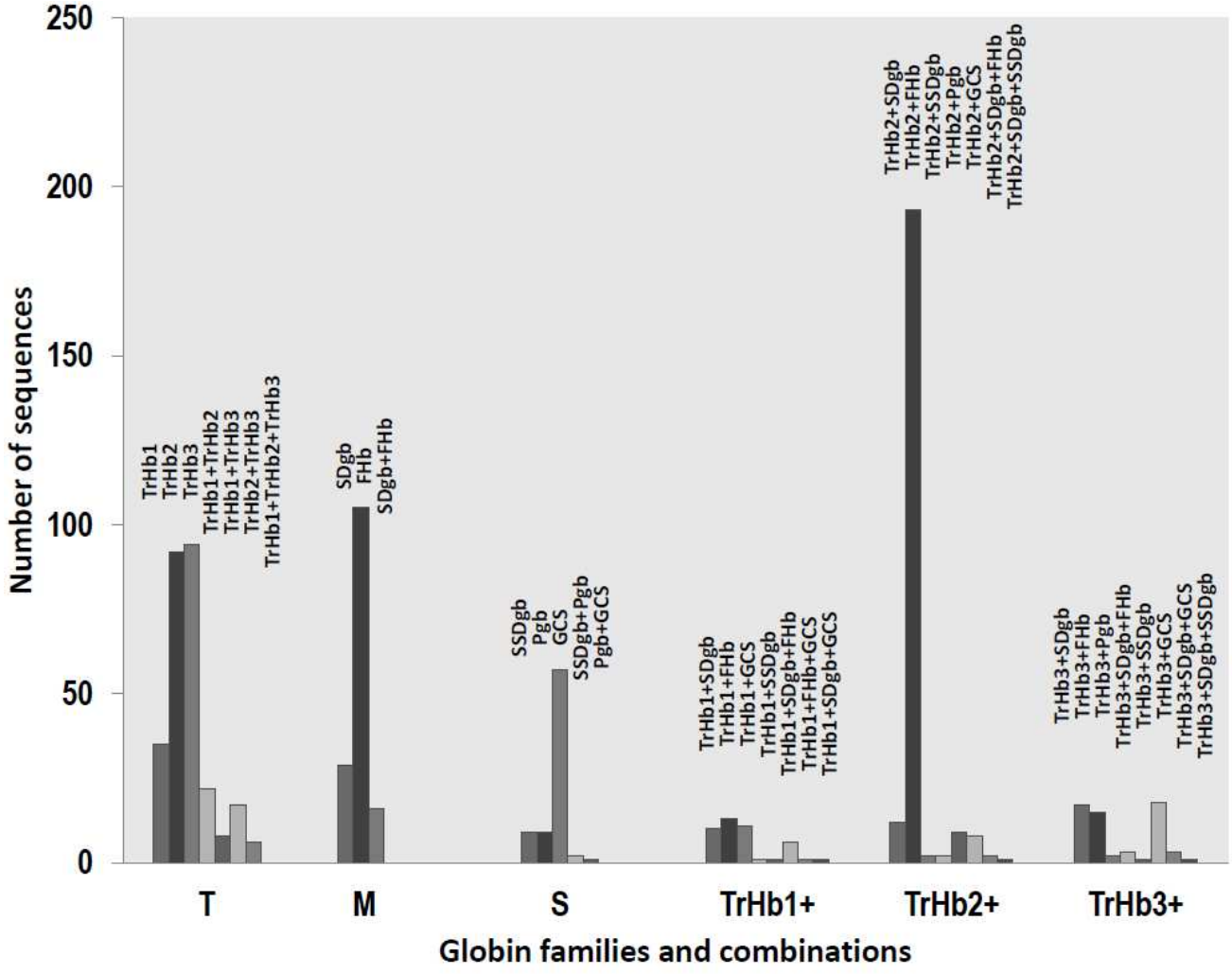


Figure 4

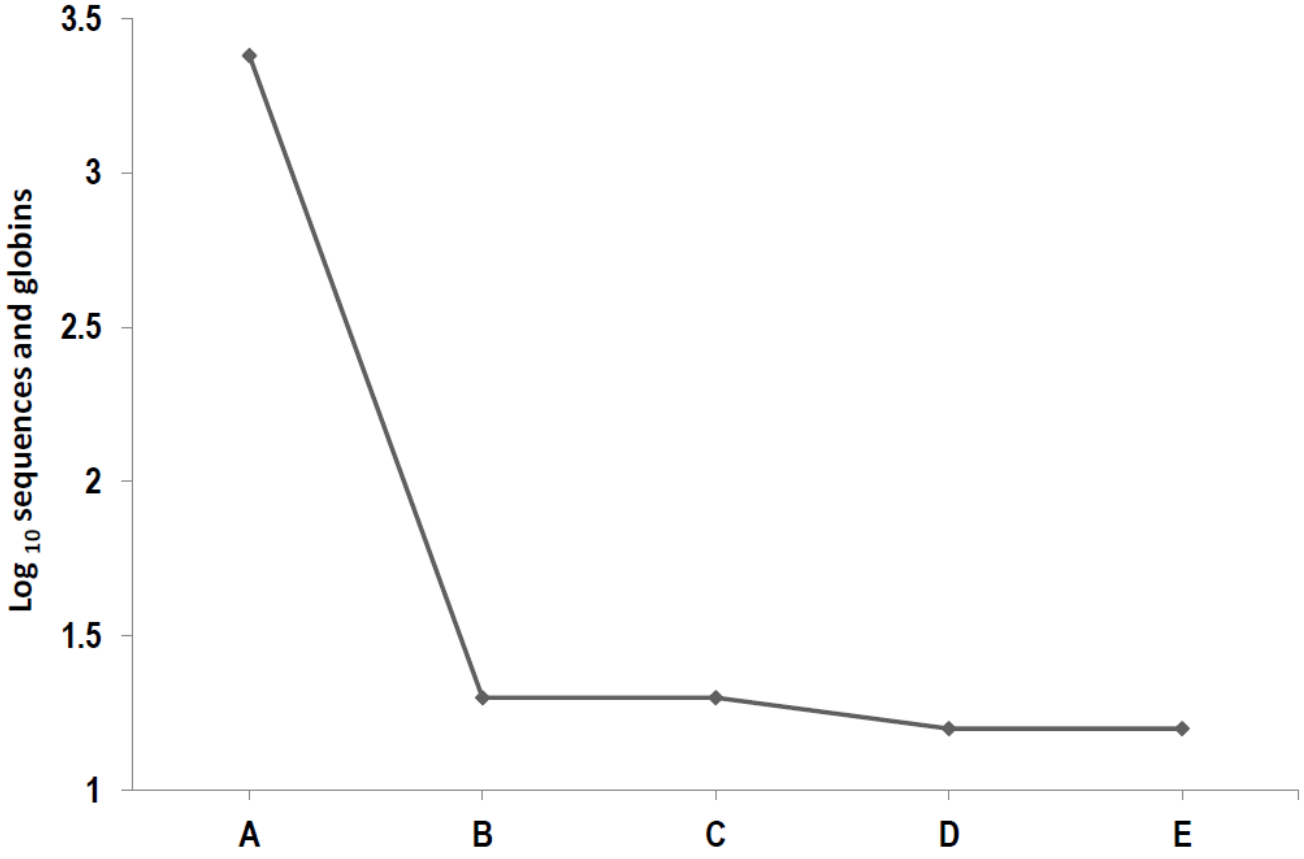


Figure 5

Highlight Manuscript Vinogradov et al

.

- Over 2200 bacterial and 140 archaeal genomes were surveyed for sequences encoding globins.
- 52% of the bacterial genomes and 32% of the archaeal genomes possess putative globins.
- Of the 147 possible combinations of the eight globin subfamilies, only 83 are observed.
- A comprehensive nomenclature involving prokaryotic and eukaryotic globins is proposed.
- We review the experimental approaches aimed at the identification of globin function in bacteria.